

Maura Quaquarelli

*Utilizzo del discovery tool dell'Università di Bologna:
un'analisi quantitativa e qualitativa*

1. Introduzione

La presente ricerca nasce con l'intento di analizzare il comportamento degli utenti e le metodologie di ricerca all'interno del discovery tool dell'Università di Bologna, denominato AlmaStart.¹ Sempre più spesso il catalogo è l'unico strumento di intermediazione fra il materiale bibliografico e i ricercatori e gli studenti,² in quanto una parte considerevole dei documenti è disponibile in formato elettronico e per ottenerla non è più necessario recarsi fisicamente in biblioteca. È dunque essenziale che il catalogo risponda in maniera adeguata alle necessità degli utenti che nella maggior parte dei casi non si interfacciano con un bibliotecario che possa fornir loro assistenza. Quindi analizzare l'uso che viene fatto di cataloghi e discovery permette di

¹ L'articolo è una rielaborazione della mia tesi magistrale del corso Scienze del libro e del documento discussa a luglio 2021 presso l'Università degli Studi di Bologna, con relatrice prof.ssa Lucia Sardo e correlatrice prof.ssa Fiammetta Sabba.

² L'articolo Galeffi 2017 presenta una riflessione teorica sulla comunicazione implicita ed esplicita che avviene fra cataloghi ed utenti.

avere un feedback su come questi vengano utilizzati e quali siano le funzionalità che potrebbero essere migliorate.

I dati sono stati analizzati mediante un approccio combinato di analisi quantitativa e qualitativa, che hanno permesso di ricavare informazioni distinte fra loro. L'analisi quantitativa è stata utilizzata per l'intero set di dati preso in considerazione e ha portato a delineare quali sono le modalità di ricerca più utilizzate e l'evolversi di queste nel tempo: utilizzo della ricerca semplice o avanzata, degli operatori booleani, della ricerca per campi specifici quali il loro titolo o autore, il numero di parole utilizzate per la ricerca e i termini più usati. L'analisi qualitativa è stata utilizzata invece per analizzare un campione limitato delle specifiche stringhe di ricerca immesse dagli utenti; per fare questo sono state stabilite giornate campione di cui analizzare le ricerche effettuate per verificare cosa venga cercato e con quali modalità. Infine sono state analizzate tutte le stringhe che non hanno portato ad alcun risultato di ricerca per verificare perché il discovery non abbia fornito risposta. In totale sono state analizzate più di 5.500 stringhe di ricerca.

Si è scelto di prendere in considerazione quattro bimestri, i primi due del 2020 e i primi due del 2021, per osservare l'evolversi nel tempo dell'uso del discovery. Questa scelta ha permesso di verificare se e come l'improvvisa chiusura delle biblioteche causata dalla pandemia di Coronavirus avvenuta nel secondo bimestre del 2020 abbia influenzato l'utilizzo del discovery tool.

La fonte dei dati analizzati è Google Analytics. Per l'analisi quantitativa sono stati utilizzati Microsoft Excel e, in maniera molto minore, Fogli di Google, mentre per l'analisi qualitativa è stato utilizzato un software appositamente sviluppato per l'analisi e categorizzazione di dati testuali: Atlas.ti. Nella letteratura biblioteconomica internazionale si trovano diversi studi sull'utilizzo del discovery tool o di cataloghi utilizzando Google Analytics come fonte dei dati,³ ma non sono

³ Si vedano ad esempio gli articoli Fischer 2021, Cohen 2015 e, sul territorio italiano, Fusetti 2019.

presenti invece analisi delle stringhe di ricerca effettuate dagli utenti mediante programmi specifici per l'analisi qualitativa di dati testuali come Atlas.ti o programmi equivalenti.

Il caso specifico analizzato è l'utilizzo di AlmaStart, il discovery dell'Università di Bologna basato su EBSCO Discovery Service (d'ora in poi EDS) di EBSCO⁴ che è stato implementato nel corso del 2017 con l'obiettivo di rendere più facilmente accessibile alla propria comunità accademica la grande varietà di risorse fisiche e digitali presenti nelle collezioni delle biblioteche.

Alle risorse attivate nell'indice centrale dello strumento, collegate al possesso elettronico dell'Università tramite il link resolver SFX, sono state aggiunte un insieme di risorse locali:⁵ l'intero Catalogo del Polo Bolognese, in cui è possibile reperire il possesso fisico delle biblioteche e gli e-book, i posseduti dell'Università e della Città Metropolitana di Bologna del catalogo ACNP, in cui sono presenti periodici cartacei e online, e le collezioni digitali AlmaDL di risorse prodotte dall'Università di Bologna.

Il discovery dispone di una propria interfaccia dalla quale è possibile effettuare una ricerca semplice oppure passare a quella avanzata. È possibile lanciare una ricerca base anche utilizzando il motore di ricerca presente sul portale delle biblioteche dell'Università e su alcuni siti di specifiche biblioteche, oppure utilizzando il tool "AlmaStart al Volo" che permette di evidenziare parole su un browser e lanciare una ricerca nel discovery.⁶

Nei paragrafi seguenti, dopo aver delineato lo scopo della ricerca, viene illustrata la metodologia di lavoro adottata, sia per quanto

⁴ A partire da gennaio 2022 il discovery dell'Università di Bologna è basato su un diverso prodotto commerciale, ma l'intera analisi del presente articolo ha come base i dati del discovery EDS.

⁵ L'illustrazione di un caso specifico di scelte e configurazioni da effettuare per l'implementazione del discovery EDS si trova in Bardelli 2019.

⁶ "AlmaStart al volo" è un preferito da aggiungere al proprio browser contenente un comando javascript che invia una ricerca al discovery.

riguarda la strutturazione e la conseguente scelta dei dati di Google Analytics, sia riguardo alla successiva analisi quantitativa e qualitativa dei dati. Vengono poi presentati i principali risultati ottenuti che delineano in maniera abbastanza chiara come e cosa cerchino gli utenti all'interno del discovery. L'analisi puntuale di un sottoinsieme di stringhe ha fatto emergere inoltre un insieme di elementi da cui è possibile partire per migliorare le funzionalità dello strumento.

Il valore aggiunto del secondo tipo di analisi, sicuramente più dispendiosa in termini di tempo rispetto alle consuete analisi quantitative, è che si è riusciti a osservare l'approccio di ricerca e a intercettare le necessità di un campione di utenti che hanno utilizzato lo strumento. Questi utenti non coincidono necessariamente con chi si reca in biblioteca o si rivolge al servizio di reference online,⁷ ma allarga la platea degli utenti a chiunque provi a utilizzare il discovery autonomamente per soddisfare una qualche necessità informativa.

2. Obiettivi e finalità

La presente ricerca si è stabilita come obiettivo di analizzare i dati d'uso del discovery dell'Università di Bologna raccolti da Google Analytics. Oltre ai dati di accesso e di utilizzo dello strumento, è stato analizzato anche come vengano effettuate le ricerche e quali siano quelle che non portano ad alcun risultato e, stabilito un campione di dati, quale ne sia l'oggetto e quali elementi contengano al loro interno.

La finalità della ricerca è stata cercare di capire come gli utenti utilizzino il discovery analizzando il trend di comportamento per verificare se questo strumento sia in grado di fornire risposte adeguate a

⁷ Altri studi qualitativi sull'utilizzo di discovery e cataloghi, come ad esempio Francese 2013 oppure Bianchini 2017, rilevano i dati di utenti che si recano fisicamente in biblioteca.

chi lo utilizza o ci siano funzionalità aggiuntive che potrebbero essere implementate.

Si è voluto inoltre verificare, come già accennato nell'introduzione, se e in che modo la chiusura delle biblioteche e il lockdown dovuto alla pandemia di coronavirus iniziato a marzo 2020 abbia impattato sulle modalità di ricerca degli utenti.

3. Metodologia e strumenti

3.1. Raccolta dati

I dati utilizzati per la presente ricerca sono stati ricavati da Google Analytics. Questo strumento di web analytics, grazie all'introduzione di script all'interno delle pagine web da monitorare, registra le pagine visualizzate e le azioni che vengono compiute da un utente su un determinato sito, anonimizzando le informazioni rilevate e non violando quindi la privacy degli utenti⁸.

Da Google Analytics sono state estrapolate le informazioni di insieme sull'utilizzo del sito nei quattro periodi analizzati: il numero di utenti che hanno visitato il sito, il numero di sessioni, di pagine visualizzate e la durata media di una sessione.

Non si è potuto ricavare alcuna informazione sugli utenti che hanno effettuato le ricerche, come ad esempio la tipologia di utente all'interno dell'Ateneo di Bologna (studente, docente, dottorando, ecc.): tali informazioni infatti sono legate all'account dell'Università con cui si può autenticare l'utente, ma non vengono esposte a sistemi esterni come Google Analytics.

Sono state poi utilizzate le informazioni relative agli specifici eventi, ovvero le ricerche e le azioni compiute dagli utenti sul discovery.

Su Google Analytics ogni evento registrato ha tre attributi che lo

⁸ Cfr. Marek 2011.

distinguono, “categoria”, “azione” ed “etichetta”: la categoria mostra la tipologia di evento e una parte di URL della pagina in cui è avvenuto, mentre azioni ed etichette variano in base alla categoria dell'evento.⁹ Nello specifico sono state analizzate le categorie presenti nella Tabella 1.

Categoria	Azione	Etichetta
Search Term	Stringa di ricerca	ID di sessione
Custom Link¹⁰	Titolo del record su cui è stato selezionato il custom link	Stringa di ricerca
Detailed Record	Tipologia di azione: Salva, Cita, Esporta...	Titolo del record su cui è stata fatta l'azione

Tabella 1 - *Principali categorie, azioni ed etichette degli eventi utilizzati per l'analisi del discovery tool EDS*

Le azioni relative alla categoria “*Search terms*” contengono al loro interno diversi elementi (tabella 2). Oltre alla stringa di ricerca immessa dall'utente si possono trovare codici di campo, operatori booleani ed eventuali filtri per i cataloghi locali. La stringa è inoltre preceduta da [Guest] quando è stata effettuata da un utente che non si è collegato da un IP appartenente alla rete di Ateneo e non si è loggato con il servizio proxy. Se l'utente si logga dopo aver effettuato la ricerca, questa viene registrata due volte, una con i termini cercati preceduti da [Guest] e una con i soli termini cercati.

⁹ Una presentazione approfondita di come vengano visualizzati i dati di EDS in Google Analytics è presente in Fischer 2021.

¹⁰ Link presenti nella lista dei risultati di ricerca e all'interno dei singoli record bibliografici. Possono portare al full-text di una risorsa, al link resolver o alla visualizzazione di un record all'interno di un altro catalogo o banca dati.

Elementi azioni relativa a Search terms	Descrizione	Esempio
Termini cercati	Parole immesse nel box di ricerca del discovery	“fonetica francese e tedesca” ¹¹
Codici di campo	Appaiono quando la ricerca è stata effettuata su un campo specifico, oppure possono essere inseriti manualmente nella stringa di ricerca	“AU Giuseppina gentili” AU effettua una ricerca nel campo autore
Operatori booleani	Possono essere inseriti manualmente dall'utente all'interno della stringa oppure inseriti automaticamente dalla maschera di ricerca quando si cerca con combinazione di campi	“chitosan AND coating AND bioplastics”
Filtri cataloghi	Compaiono quando l'utente utilizza i filtri relativi alle collezioni locali nella ricerca avanzata oppure nei filtri presenti a lato dei risultati di ricerca	“(matlab) AND LN cat05251a” LN cat05251a è il filtro che indica il Catalogo del Polo Bolognese
Guest	Se l'utente non si collega da un indirizzo IP appartenente all'Università di Bologna e non si autentica con il servizio proxy viene considerato guest, utente ospite.	“[Guest] agv handling system”

Tabella 2 - *Elementi presenti nelle azioni relative alla categoria Search Terms di Google Analytics*

All'interno dei *Search terms* appaiono sia le ricerche effettuate con la ricerca semplice sia quelle effettuate tramite la ricerca avanzata; queste si possono distinguere osservando l'intera denominazione del-

¹¹ Tutti gli esempi riportati fra virgolette sono tratti dalle stringhe di ricerca analizzate.

la categoria evento che riporta parte della URL della pagina in cui è avvenuto l'evento. Sempre dalla URL si possono vedere le ricerche per cui è stato visualizzato un record nel dettaglio; viene registrato un evento per ogni record visualizzato, con la relativa azione che riporta la ricerca effettuata. Vengono infine distinte le ricerche che non hanno portato ad alcun risultato, identificate dalla parte di categoria "No results".

Categoria evento	Azione evento	Descrizione
EBSCO:Search Term /eds/ results	"[Guest] vite in cantiere"	Ricerca semplice
EBSCO:Search Term /eds/ resultsadvanced	"[Guest] bologna AND TI strage"	Ricerca avanzata
EBSCO:Search Term /eds/detail/ detail	"videogames climate change"	Ricerca per cui è stato visualizzato un record nel dettaglio ¹²
EBSCO:Search Term [No Results]/eds/search/basic	"Adler, E. (1997) "Seizing the Middle Ground: Constructivism in World Politics." European Journal of International Relations 3.3: 319-363."	Ricerca semplice che non ha portato ad alcun risultato

Tabella 3 - Esempi di categorie e azioni evento di Google Analytics

Per estrapolare i dati da Google Analytics sono stati creati report personalizzati che riportano Categoria/Evento/Etichetta delle azioni registrate negli intervalli di tempo presi in considerazione.

¹² Il singolo titolo visualizzato viene invece registrato con l'azione *Detailed Record*.

3.2. Lavorazione dati

I dati estratti da Google Analytics sono stati elaborati mediante l'utilizzo di 3 software: Microsoft Excel, Google Fogli e Atlas.ti.

Mediante Microsoft Excel è stata fatta tutta l'analisi quantitativa dei dati e delle stringhe estrapolate. Con lo stesso software sono state effettuate operazioni di data cleaning (ad esempio l'eliminazione della parte “[Guest]” dalla stringa di ricerca) e l'individuazione e il conteggio di specifiche parti della stringa (ad esempio gli operatori booleani o l'utilizzo di codici di campo).

È opportuno specificare che i conteggi effettuati su operatori booleani e codici di campo potrebbero essere leggermente sovrastimati: infatti nel caso l'utente abbia digitato un titolo o una chiave di ricerca con tutte le parole maiuscole, eventuali parole coincidenti con operatori booleani o codici di campo sono stati conteggiati come tali. Ad esempio, nella stringa “TI “POLITICS AND GOVERNANCE IN THE UK.”” è stato conteggiato un codice di campo TI, titolo, ma anche un operatore booleano AND, che è contenuto del titolo cercato. Lo stesso discorso vale per i caratteri speciali asterisco e virgolette, che occasionalmente potrebbero essere presenti nelle stringhe anche come semplici parti del testo cercato.

L'applicazione Fogli di Google è stata utilizzata esclusivamente per individuare automaticamente la lingua usata nelle stringhe di ricerca poiché dispone di una funzione apposita.

Il software Atlas.ti, sviluppato per eseguire analisi qualitative sui dati, è stato utilizzato per l'analisi qualitativa di un sottoinsieme delle stringhe di ricerca immesse dagli utenti nel discovery.¹³ All'interno di questo

¹³ Nella prima fase della ricerca si è valutato se fosse più opportuno utilizzare il software open source IRaMuTeQ, anch'esso indirizzato all'analisi qualitativa di dati. La necessità di pulire preventivamente i dati da analizzare, eliminando caratteri significativi per questa ricerca come gli asterischi e le virgolette, e la maggiore complessità per l'installazione e l'utilizzo di IRaMuTeQ, hanno fatto ricadere la scelta su Atlas.ti. Quest'ultimo è infatti caratterizzato dall'essere particolarmente

software sono stati importati gli insiemi delle stringhe di ricerca effettuate nelle giornate analizzate e, in un progetto separato, tutte le stringhe di ricerca che non hanno prodotto alcun risultato nel discovery.

Dopo l'importazione si è iniziato a codificare immediatamente le ricerche, attribuendo uno o più codici a ogni stringa presente. L'analisi è iniziata con procedure di codifica bottom-up, partendo dai dati e creando le categorie. Non si aveva infatti chiaro a priori quali categorie attribuire o quali tipo di informazioni codificare, ma si è avviato il lavoro rilevando tutto quanto potesse essere utile a capire che tipo di ricerca fosse stata fatta: l'utente stava cercando di localizzare o raggiungere un documento di cui conosceva a priori l'esistenza oppure stava cercando documenti su un determinato argomento? Nel primo caso, che elementi bibliografici utilizzava per la ricerca? Nel secondo, invece, in quale lingua effettuava la ricerca?

Per le stringhe che non hanno prodotto risultati, si è cercato di capire che tipologia di ricerca stesse facendo l'utente, se per documento conosciuto o argomento, e perché la ricerca non abbia prodotto alcun risultato.

Stringa di ricerca	Codici attribuiti
10435	1.Non determinato
9788869741326	1.Documento conosciuto: Libro r.ISBN tr:ISBN;1
“Il matrimonio canonico tra principi astratti e casi concreti”	1.Documento conosciuto: Libro r.Titolo re.Virgolette tr:Titolo;1

te user-friendly e di disporre di un'interfaccia molto semplice e intuitiva (Milesi 2011): è stato sufficiente visualizzare i video presenti sul sito del produttore per poter iniziare a lavorare operativamente sui dati, approfondendo in un secondo momento la lettura delle parti del manuale sulle funzioni di interesse.

AA.VV., Lineamenti di parte speciale - Ed. Monduzzi - Bologna	1.Documento conosciuto: Libro r.Autore r.Editore r.Luogo r.Titolo tr:Titolo_Autore_Luogo_Editore;4 z:Citazione bibliografica
agv handling system	1.Argomento l.Inglese
agv movimentazione materiali	1.Argomento l.Italiano
ai wei wei	1.Nome
al cardinal di trento	1.Argomento l.Italiano
al carinal di trento	1.Argomento l.Italiano
algebra	1.Argomento l.Italiano
AU andrea dell'anguillara	1.Nome re.Codice campo

Tabella 4 - Esempi di stringhe di ricerca e relativi codici inseriti in Atlas.ti

Lavorare con Atlas.ti ha permesso di creare l'intero sistema di codifica delle informazioni gradualmente, dando la possibilità di modificare, fondere o dividere codici già attribuiti. L'organizzazione di questi codici all'interno del programma non vincola a rispettare gerarchie o strutture predefinite, pertanto può essere creata man mano raggruppando i codici in insiemi omogenei.

Per farlo generalmente si antepone ai codici una sigla, in modo da visualizzare in sequenza quelli che trattano lo stesso tipo di informazione.¹⁴ È inoltre possibile suddividere i codici in gruppi, che non creano una vera e propria gerarchia ma che servono per visualizzare a posteriori par-

¹⁴ Cfr. Friese 2020.

te dei dati tramite alcune funzioni messe a disposizione dal programma.

La categorizzazione di tutte le stringhe ha in primo luogo cercato di identificare il contenuto della ricerca; in particolare si è cercato di capire se l'utente stesse facendo una ricerca esplorativa per argomento oppure cercasse di trovare un oggetto conosciuto; fra questa seconda tipologia sono stati identificati i singoli documenti bibliografici (libri, articoli, ecc.), e gli strumenti che permettono di trovare documenti bibliografici: banche dati e cataloghi¹⁵.

Fra le tipologie non identificate ci sono le stringhe contenenti solamente un nome proprio, perché non c'era modo di sapere se quello fosse inteso come autore o soggetto del documento desiderato, le ricerche in cui c'erano molteplici possibilità e quelle per le quali non era possibile determinare il contenuto.

Per identificare l'oggetto della ricerca, le stringhe sono state cercate in Google, in Google Scholar e quando necessario in ACNP, nel Catalogo del Polo Bolognese o in banche dati specifiche. Non è stato appositamente utilizzato AlmaStart, perché gran parte dell'analisi è stata effettuata durante l'ultimo bimestre analizzato e non si voleva influenzare i dati di ricerca del discovery che sarebbero stati analizzati in seguito.

Per le ricerche dei documenti conosciuti, oltre a identificare la specifica tipologia del documento (come libro, articolo...), si è rilevato con quali elementi bibliografici è stato cercato, la combinazione di questi e se la ricerca fosse stata fatta mediante una citazione bibliografica. Per le ricerche per argomento si è indicata la lingua di ricerca. Infine, se alcune stringhe contenevano elementi di ricerca generalmente

¹⁵ Negli studi sul comportamento degli utenti sui cataloghi di biblioteca generalmente si distinguono le ricerche di oggetti conosciuti e quelle per argomento, anche dette esplorative (cfr. ad esempio Bianchini 2020). Nell'analisi delle stringhe e nelle successive tabelle che riportano i risultati è sembrato utile distinguere immediatamente le ricerche di oggetti conosciuti in diverse categorie: le ricerche di documenti noti all'utente e le ricerche di altri strumenti di ricerca, le banche dati e i cataloghi. Infatti, trattandosi dell'analisi di un discovery tool, uno strumento che aggrega dati, si voleva capire se l'utente si aspettasse di trovare indicizzate anche le banche dati e gli altri strumenti di ricerca messi a disposizione.

non ricercabili all'interno di un discovery, sono state segnalate.

Per le ricerche senza alcun risultato, dopo aver identificato l'oggetto della ricerca, si è cercato invece di capire il perché: i dati bibliografici erano scorretti? C'erano elementi che il discovery non era in grado di interpretare correttamente?

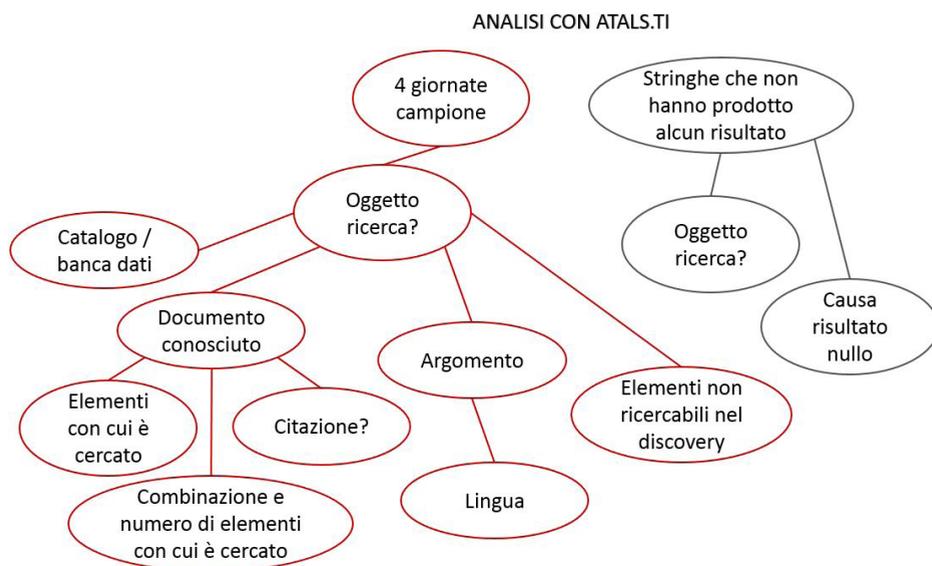


Figura 1 - Schema elementi codificati con *Atlas.ti*

Al termine del lavoro sono emersi 139 codici appartenenti a 6 gruppi, a cui si aggiunge il codice “Citazione bibliografica” non appartenente ad alcun gruppo (tabella 5).

Denominazione gruppo	Numero codici contenuti	Descrizione
Tipologia di ricerca	22 codici	Tipologia di ricerca che effettua l'utente sul discovery: 1.Argomento 1.Banca dati 1.Documento conosciuto: Libro...

Lingue ricerche argomento	13 codici	Lingua in cui viene effettuata la ricerca per argomento: l.Italiano l.Inglese l.Cinese...
Elementi ricerca per documento conosciuto - singoli	24 codici	Elementi di cui è composta la stringa di ricerca di un documento conosciuto: r.titolo r.autore r.data...
Elementi ricerca per documento conosciuto - raggruppati	52 codici	Insieme di elementi di cui è composta la stringa di ricerca di un documento conosciuto e numero di elementi contenuti: tr:Titolo_Autore;2, tr:Titolo_Autore_Edizione_Luogo_Editore_Data;6 tr:Titolo_Autore_Nome Monografia;3...
Non da cercare nel discovery	12 codici	Stringhe che fanno presupporre che l'utente stia cercando qualcosa non contenuto nel discovery: ndt.Corso accademico ndt.Collocazione ndt.Traduzione di termine...
Possibile causa risultato nullo	16 codici	Possibile causa del perché la ricerca non abbia prodotto alcun risultato all'interno nel discovery: pr:Battitura: spelling pr:Citazione pr:Documento non indicizzato

Tabella 5 - *Gruppi di codici creati in Atlas.ti*

Con Atlas.ti è stato inoltre possibile verificare la frequenza di tutti i termini italiani e inglesi presenti nelle stringhe di ricerca degli interi bimestri analizzati. Infatti sono disponibili funzioni che permettono di creare una lista con la frequenza delle parole inserite e le *world cloud*.

4. Risultati

4.1. Analisi quantitativa dati bimestrali

L'analisi quantitativa dei dati dei primi due bimestri del 2020 e 2021 ha permesso di delineare le modalità di ricerca degli utenti e l'evoluzione nel tempo di queste, in particolare in relazione alla chiusura completa delle biblioteche avvenuta nel secondo bimestre del 2020¹⁶. Oltre ai dati generali di utilizzo del sito, sono state esaminate le stringhe di ricerca e le azioni principali che avvengono quando appare l'elenco dei risultati.

Partendo dall'impatto che la chiusura delle biblioteche ha avuto sull'uso di AlmaStart si nota che la variazione più evidente ha riguardato il numero di utenti, sessioni e ricerche di chi lo ha utilizzato. Infatti i dati del secondo bimestre 2020 sono aumentati tutti di oltre il 79% rispetto al bimestre precedente. Tale incremento si registra anche fra il primo e il secondo bimestre dell'anno successivo, ma è compreso fra il 25% e il 29% (tabella 6). I dati relativi al 2019, anno precedente alla mia analisi, mostrano che l'utilizzo del discovery era già in crescita, a parte un calo durante i mesi estivi, ma che l'incremento del secondo bimestre del 2020 è nettamente più marcato (figura 2): da questo si evince che il maggiore utilizzo sia relazionato alla chiusura delle biblioteche.

La maggior parte degli altri dati raccolti, a partire dalle pagine visualizzate per sessione e la durata di questa, non registrano invece grosse variazioni, anzi in molti casi i dati del secondo bimestre del 2020 sono assolutamente in linea con gli altri periodi analizzati.

Ogni utente visualizza in media quasi 5 pagine per sessione e resta sul discovery per più di 6 minuti, dati che fanno presupporre che gli utenti trovino qualcosa di interessante per le loro ricerche e che lo utilizzino per un po' di tempo prima di spostarsi su altri siti (tabella 6).

¹⁶ L'Università di Bologna ha sospeso tutte le attività didattiche dal 24 al 29 febbraio 2020 in seguito all'entrata in zona rossa della Regione Emilia-Romagna; queste sono state poi riprese in modalità online a partire dal 3 marzo 2020.

	gen-feb 2020	mar-apr 2020	Aumento fra 1° e 2° bimestre 2020	gen-feb 2021	mar-apr 2021	Aumento fra 1° e 2° bimestre 2021
Utenti	8.355	15.124	81,02%	11.493	14.800	28,77%
Sessioni	15.623	29.398	88,17%	23.972	29.979	25,06%
Ricerche	34.033	60.934	79,04%	50.368	63.038	25,15%
Visualizzazio- ni di pagina	77.987	143.308	83,76%	115.798	148.165	27,95%
Pagine/ses- sione	4,99	4,87	-2,40%	4,83	4,94	2,28%
Durata sessio- ne media	00:06:13	00:06:32	5,09%	00:06:26	00:06:39	3,37%

Tabella 6 - *Dati generali di utilizzo del discovery nei quattro bimestri analizzati e percentuale di aumento fra il primo e il secondo bimestre*

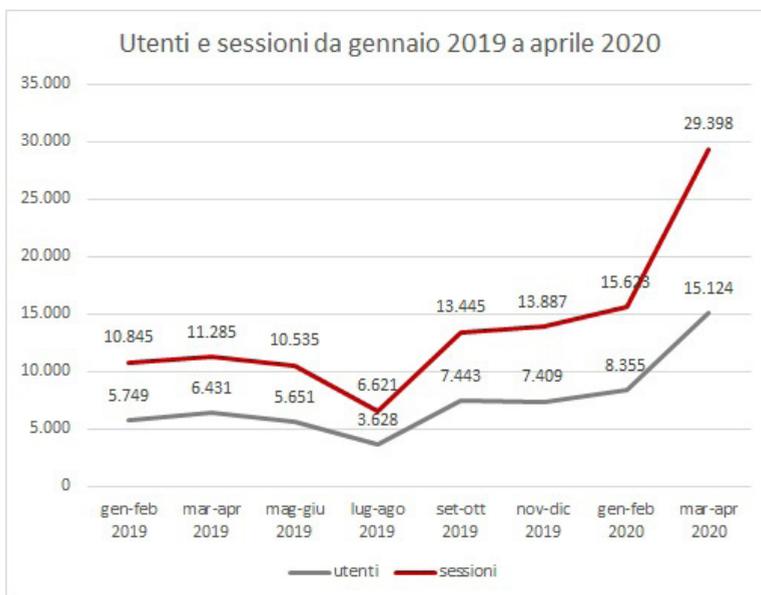


Figura 2 - *Grafico che raffigura il numero di utenti e delle sessioni sul discovery fra gennaio 2019 e aprile 2020*

Le ricerche effettuate da utenti Guest che si collegano al discovery all'esterno della rete Unibo e che non si autenticano mediante il proxy restano inferiori al 44% anche nel secondo bimestre 2020 quando l'Università era chiusa e i collegamenti avvenivano quindi necessariamente da rete esterna all'Ateneo (tabella 7). Questo dato fa riflettere sul fatto che l'utente all'interno del discovery riesce ad autenticarsi facilmente ed è spinto a farlo perché l'accesso alle risorse a testo pieno a pagamento è possibile solo previa autenticazione.

	gen-feb 2020	%	mar-apr 2020	%	gen-feb 2021	%	mar-apr 2021	%
Unibo	21.368	62,79%	34.136	56,02%	31.881	63,30%	43.128	68,42%
Guest	12.665	37,21%	26.798	43,98%	18.487	36,70%	11.910	31,58%
TOTALE	34.033		60.934		50.368		63.038	

Tabella 7 - Numero di ricerche divise fra utenti Unibo e Guest nei quattro bimestri analizzati

Per quanto concerne le modalità di ricerca, la ricerca base è utilizzata per più del 92,5% dei casi in ogni bimestre (tabella 8). Il dato è sicuramente in parte influenzato dal fatto che tramite il motore di ricerca presente nel sito delle biblioteche e mediante il tool "AlmaStart al volo" sia possibile lanciare solo ricerche semplici.

Si suppone che lo scarso utilizzo della ricerca avanzata possa dipendere dal fatto che da parte degli utenti sia l'unica modalità di ricerca conosciuta, la stessa usata nei motori di ricerca. Ma dato che probabilmente viene utilizzato anche da utenti in grado di formulare ricerche più articolate, può essere indizio del fatto che già tramite la ricerca base sia possibile ottenere i risultati desiderati e quindi non sia sentita la necessità di fare ricerche più specifiche.

	gen-feb 2020	%	mar-apr 2020	%	gen-feb 2021	%	mar-apr 2021	%
Ricerca base	32.141	94,44%	56.524	92,76%	48.174	95,64%	58.621	92,99%

Ricerca avanzata	1.892	5,56%	4.410	7,24%	2.194	4,36%	4.417	7,01%
TOTALE	34.033		60.934		50.368		63.038	

Tabella 8 - Numero di ricerche base e avanzate nei quattro bimestri analizzati

Gli operatori booleani, i caratteri speciali e i codici di ricerca per campi specifici possono essere inseriti sia manualmente dall'utente che in automatico dal discovery nel momento in cui si seleziona una combinazione di campi, un campo specifico in cui cercare oppure si clicca su un elemento di un record bibliografico per estendere la ricerca ad altri record con lo stesso campo (ad esempio cliccando sul soggetto "SENTIMENT analysis" viene lanciata la ricerca "DE "SENTIMENT analysis" dove DE indica il campo soggetto).

Il totale delle ricerche che contengono gli operatori booleani, i caratteri speciali e delle ricerche mediante campi specifici non supera l'11%; gli elementi più utilizzati sono l'operatore booleano AND e la ricerca per campi titolo o autore (tabelle 9 e 10).

	gen-feb 2020	% sul totale ricerche	mar-apr 2020	% sul totale ricerche	gen-feb 2021	% sul totale ricerche	mar-apr 2021	% sul totale ricerche
AND	885	2,60%	3.104	5,09%	1.272	2,52%	3.146	4,99%
OR	141	0,41%	328	0,53%	160	0,31%	678	1,07%
NOT	1	0,01%	48	0,07%	29	0,05%	38	0,06%
*	140	0,41%	194	0,31%	223	0,44%	553	0,87%
“	848	2,49%	1.547	2,53%	1.550	3,07%	1.947	3,08%
TOTALE	2.015	5,92%	5.221	8,56%	3.234	6,42%	6.362	10,09%

Tabella 9 - Ricerche con operatori booleani e caratteri speciali e percentuale rispetto alle ricerche totali nei quattro bimestri analizzati

Codice	gen-feb 2020	% sul totale ricerche	mar-apr 2020	% sul totale ricerche	gen-feb 2021	% sul totale ricerche	mar-apr 2021	% sul totale ricerche
TI – Ti- tolo	1.199	2,38%	2.934	4,81%	2.050	4,07%	2080	3,30%
AR o AU – Autore	1.152	2,28%	2.379	3,90%	1.597	3,17%	1.663	2,63%
DE o SU –Soggetto	279	0,55%	534	0,87%	445	0,88%	614	0,97%
TX - Tut- ti i campi	125	0,24%	130	0,21%	74	0,14%	246	0,39%
AN - Identifi- cativo	119	0,23%	174	0,28%	80	0,15%	110	0,17%
Altri campi	200	0,39%	308	0,50%	312	0,61%	303	0,48%
TOTALE	3.074	6,10%	6.459	10,60%	4.558	9,04%	5.016	7,95%

Tabella 10 - *Ricerche per campi specifici e percentuale rispetto alle ricerche totali nei quattro bimestri analizzati*

Il numero di parole calcolate nelle stringhe di ricerca è stato conteggiato dopo averle ripulite dai filtri dei cataloghi, dai codici di ricerca per campo e dagli operatori booleani. Si notano alcune ricerche molto lunghe, che utilizzano anche più di 100 parole; queste sono generate automaticamente dal discovery nel momento in cui si utilizza la funzione “Trova risultati simili” presente nel dettaglio del record che effettua una ricerca combinata tra titolo e abstract della risorsa di partenza.

I dati ottenuti sono molti simili fra un bimestre e l'altro, infatti si evince che la maggior parte delle ricerche viene fatta utilizzando solo 2 parole e che circa il 65% delle ricerche contiene da 1 a 4 termini (tabella 11).

	gen-feb 2020	mar-apr 2020	gen-feb 2021	mar-apr 2021
Range	165 (1-165)	314 (1-314)	306 (1-306)	238 (1-238)

Media	4,35	4,21	3,93	3,83
Mediana	3	3	3	3
Moda	2	2	2	2

Tabella 11 - *Principali dati statistici relativi al numero di parole nelle stringhe di ricerca nei quattro bimestri analizzati*

Per quanto riguarda le ricerche effettuate, è stata analizzata infine la frequenza delle parole nelle stringhe di ricerca in italiano ed in inglese. Per identificare la lingua delle stringhe è stata utilizzata la funzione DETECTLANGUAGE disponibile nell'applicazione Fogli di Google;¹⁷ queste sono poi state inserite nel programma Atlas.ti, dove è possibile utilizzare stop word list personalizzabili per escludere parole non significative.¹⁸

Sia in italiano che in inglese tutti i termini più cercati afferiscono all'ambito delle scienze umane e sociali, non sono presenti invece termini relativi alle scienze pure (figure 3 e 4). Si è ipotizzato che questo comportamento possa derivare da due fattori distinti: da un lato le scienze umane e sociali presentano un maggior grado di interdisciplinarietà e quindi il discovery, che interroga simultaneamente molti database diversi, permette di fare un'unica ricerca trasversale su questi senza essere costretti a replicare la ricerca su diversi database. Inoltre, i database disciplinari relativi alle scienze pure, come la fisica, la chi-

¹⁷ La funzione non identifica la lingua con precisione nel 100% dei casi, anche perché ci sono stringhe di ricerca con termini in lingue distinte. È comunque servita a identificare con una buona approssimazione le stringhe di ricerca effettuate in italiano e in inglese. Ne sono state identificate 80.171 in italiano (38,47% delle ricerche totali) e 95.285 in inglese (45,73% delle ricerche totali). Il restante 15,8% delle ricerche è in altre lingue e nessuna di queste ha una frequenza significativa rispetto alle restanti.

¹⁸ Per l'italiano è stato necessario sostituire preventivamente tutti i caratteri apostrofo con un apostrofo seguito da spazio, perché diversamente Atlas.ti avrebbe letto la parola preceduta dall'apostrofo come una parola unica, conteggiandola separatamente dalla stessa parola non preceduta dall'apostrofo (es: "l'economia" e "economia" sarebbero state contate come due parole distinte).

è sicuramente influenzato dal fatto che quei link sono gli unici che compaiono su tutti i record presenti, mentre gli altri appaiono solo su specifici record.

In generale i custom links vengono utilizzati circa il 35% delle volte rispetto alle ricerche effettuate, a eccezione del secondo bimestre del 2020 in cui il dato è più alto (tabella 12). Questo incremento può essere dovuto dal fatto che le risorse online erano le uniche che si potevano ottenere, data la chiusura delle biblioteche, quindi gli utenti erano obbligati ad ottenere una versione online del documento ricercato.

	gen-feb 2020	% ri- spetto a ricerche totali	mar-apr 2020	% ri- spetto a ricerche totali	gen-feb 2021	% ri- spetto a ricerche totali	mar-apr 2021	% ri- spetto a ricerche totali
Accedi al full-text	1.880	5,52%	3.534	5,80%	3.027	6,01%	4.028	6,39%
Vedi il record su altra risorsa	2.056	6,04%	3.450	5,66%	2.780	5,51%	2.834	4,49%
Link resolver	7.743	22,75%	21.379	35,08%	12.572	24,96%	15.986	25,35%
Non determinato	85	0,25%	176	0,28%	103	0,20%	111	0,17%
TOTALE	11.764	34,56%	28.363	46,54%	18.379	36,48%	22.959	36,42%

Tabella 12 - *Tipologia di custom links utilizzati e percentuale rispetto alle ricerche totali nei quattro bimestri analizzati*

La percentuale di record visualizzati rimane inferiore al 35% (tabella 13); nel leggere questo dato bisogna tener conto del fatto che nella schermata dei risultati sono già disponibili i custom links e le informazioni sulle biblioteche che posseggono il documento per le risorse cartacee, pertanto l'utente non è obbligato a visualizzare il record completo per raggiungere il full text o capire dove reperirlo. Il valore più basso si registra nel secondo semestre del 2020, dato che evidenzia ancora una volta la necessità per gli utenti di raggiungere le risorse piuttosto che di cercare informazioni bibliografiche.

Meno del 13% dei record visualizzati conduce all'utilizzo di una qualche azione (tabella 14). "Cita", "Esporta" e "Stampa" sono le più utilizzate (tabella 15). Questi dati rafforzano l'idea che il discovery sia stato usato soprattutto per raggiungere le risorse e non per cercare informazioni bibliografiche, guardando ed utilizzando nel dettaglio i record trovati.

	gen-feb 2020	% ri- spetto a ricerche totali	mar-apr 2020	% ri- spetto a ricerche totali	gen-feb 2021	% ri- spetto a ricerche totali	mar-apr 2021	% ri- spetto a ricerche totali
Record di cui è stato visionato il dettaglio	11.034	32,42%	17.561	28,82%	17.147	34,04%	21.586	34,24%

Tabella 13 - Numero record di cui è stato visualizzato il dettaglio e percentuale rispetto ricerche totali nei quattro bimestri analizzati

	gen-feb 2020	% su record visualiz- zati nel detta- glio	mar-apr 2020	% su record visualiz- zati nel detta- glio	gen-feb 2021	% su record visualiz- zati nel detta- glio	mar-apr 2021	% su record visualiz- zati nel detta- glio
Azioni eseguite dal dettaglio record	896	8,12%	2.240	12,76%	1.529	8,92%	2.445	11,33%

Tabella 14 - Azioni eseguite dal dettaglio record e percentuale rispetto ai record visualizzati nei quattro bimestri analizzati

	gen-feb 2020	mar-apr 2020	gen-feb 2021	mar-apr 2021	TOTALE
Aggiungi alla cartella	43	44	34	54	175
Ascolta	0	0	2	0	2

Cita	236	473	473	740	1.922
Crea nota	5	10	6	3	24
E-mail	28	45	31	37	141
Esporta	147	452	227	363	1.189
Google Drive	65	288	135	194	682
Permalink	37	76	87	158	358
Rimuovi dalla cartella	15	47	35	65	162
Salva	206	592	361	552	1.711
Stampa	114	209	137	278	738
Traduci	0	4	1	1	6
TOTALE	896	2.240	1.529	2.445	7.110

Tabella 15 - Azioni eseguite dal dettaglio record nei quattro bimestri analizzati

4.2. Analisi qualitativa stringhe di ricerca

Mediante l'utilizzo del software Atlas.ti è stato analizzato nel dettaglio un sottoinsieme delle ricerche effettuate. Sono state scelte 4 giornate campione dei bimestri precedentemente analizzati. Nello specifico si sono scelti 4 lunedì centrali nei bimestri analizzati, dato che di solito l'utilizzo del discovery ha una deflessione nel weekend e risale fra il lunedì e il venerdì (figura 5).



Figura 5 - Numero di sessioni giornaliere fra il 1 gennaio 2020 e il 28 febbraio 2020

Prima di procedere con la categorizzazione delle stringhe sono state analizzate due giornate consecutive, il 3 e il 4 febbraio 2020, per verificare che la distribuzione della tipologia delle ricerche effettuate non variasse in maniera consistente e che quindi la scelta di una giornata specifica non influisse in maniera rilevante sui risultati. I risultati hanno mostrato che le tipologie di ricerche sono molto simili fra loro (figura 6).

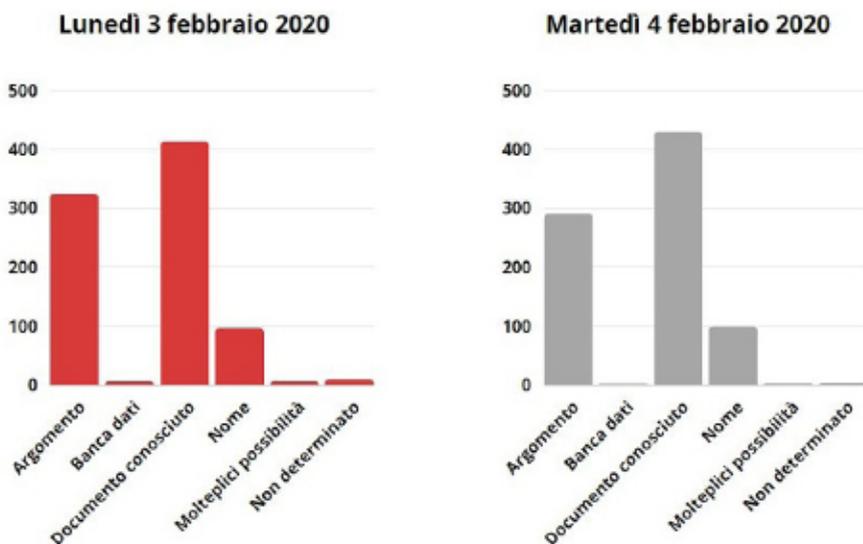


Figura 6 - Tipologie di ricerche effettuate il 3 e il 4 febbraio 2020

Oltre ai dati relativi alle giornate specifiche, sono state esaminate tutte le stringhe che non hanno prodotto alcun risultato di ricerca nel discovery nei quattro bimestri con l'intento di comprendere perché la ricerca non abbia prodotto risultati.

Il primo dato rilevato nelle 4 giornate campione riguarda l'oggetto della ricerca, per cercare di identificare cosa stesse cercando l'utente, se un documento di cui conosceva già l'esistenza, se volesse fare una ricerca per argomento oppure trovare un altro strumento di ricerca. Si è rilevato che la maggior parte delle ricerche è relativa a documenti

conosciuti, mentre le ricerche per argomento sono inferiori in ogni giornata presa in considerazione. Le ricerche di altri cataloghi sono residuali, mentre crescono nel tempo quelle relative a banche dati.

Nella seconda giornata analizzata si registra un incremento percentuale di ricerche su documenti già conosciuti dall'utente, probabilmente dovuta alla necessità di reperire copie digitali dei documenti durante la chiusura delle biblioteche. Nelle altre tre giornate si nota invece un progressivo aumento delle ricerche per argomento (tabella 16).

In tutte le giornate la tipologia di documento più cercata è quella del libro, seguito dalla ricerca di articoli e, in maniera nettamente minore, di capitoli e di periodici (tabella 17).

	Giorno 1: 3 feb 20	%	Giorno 2: 30 mar 20	%	Giorno 3: 1 feb 21	%	Giorno 4: 29 mar 21	%
Argo- mento	324	37,94%	412	28,10%	415	38,46%	457	41,58%
Banca dati	6	0,70%	11	0,75%	15	1,39%	48	4,37%
Catalogo	0	0,00%	0	0,00%	2	0,19%	1	0,09%
Docu- mento cono- sciuto	413	48,36%	878	59,89%	529	49,03%	475	43,22%
Nome	96	11,24%	146	9,96%	108	10,01%	97	8,83%
Molteplici possi- bilità	6	0,70%	5	0,34%	6	0,56%	16	1,46%
Non determi- nato	9	1,05%	14	0,95%	4	0,37%	5	0,45%
TOTALE	854		1.466		1.079		1.099	

Tabella 16 - Oggetto della ricerca effettuata nelle quattro giornate analizzate

	Giorno 1: 3 feb 20	% ri- spetto alle ricerche analiz- zate	Giorno 2: 30 mar 20	% ri- spetto alle ricerche analiz- zate	Giorno 3: 1 feb 21	% ri- spetto alle ricerche analiz- zate	Giorno 4: 29 mar 21	% ri- spetto alle ricerche analiz- zate
Articolo	122	14,29%	220	15,01%	124	11,49%	117	10,65%
Capitolo	13	1,52%	43	2,93%	26	2,41%	20	1,82%
CD- ROM	0	0,00%	0	0,00%	2	0,19%	0	0,00%
Collana	0	0,00%	0	0,00%	8	0,74%	0	0,00%
Film	3	0,35%	0	0,00%	0	0,00%	0	0,00%
Libro	254	29,74%	553	37,72%	320	29,66%	272	24,75%
Locandi- na	1	0,12%	0	0,00%	0	0,00%	0	0,00%
Mano- scritto	0	0,00%	1	0,07%	0	0,00%	0	0,00%
Norma tecnica	1	0,12%	1	0,07%	5	0,46%	4	0,36%
Numero mono- grafico periodico	0	0,00%	0	0,00%	1	0,09%	0	0,00%
Periodi- co	17	1,99%	58	3,96%	43	3,99%	60	5,46%
Sentenza	0	0,00%	0	0,00%	0	0,00%	1	0,09%
Moltepli- ci possi- bilità	2	0,23%	0	0,00%	0	0,00%	1	0,09%
Non de- terminato	0	0,00%	2	0,14%	0	0,00%	0	0,00%
TOTALE	413	48,36%	878	59,89%	529	49,03%	475	43,22%

Tabella 17 - *Tipologia di documento conosciuto cercato e percentuale rispetto a tutte le ricerche della giornata nelle quattro giornate analizzate*

Per quanto riguarda i documenti conosciuti, è emerso che il titolo e l'autore sono gli elementi più utilizzati per la ricerca, dato abbastanza scontato in quanto sono gli elementi che si ricordano di solito di un documento. Sono stati utilizzati anche altri 22 elementi, con una cinquantina di combinazioni possibili, ma nessuno di questi spicca in maniera evidente (tabella 18).

	Giorno 1: 3 feb 20	% rispetto a documenti cercati	Giorno 2: 30 mar 20	% rispetto a documenti cercati	Giorno 3: 1 feb 21	% rispetto a documenti cercati	Giorno 4: 29 mar 21	% rispetto a documenti cercati
Autore	96	23,24%	234	26,65%	136	25,71%	91	19,16%
Autore_monografia (per ricerca capitoli)	2	0,48%	2	0,23%	0	0,00%	1	0,21%
Banca dati	0	0,00%	0	0,00%	1	0,19%	0	0,00%
Codice_PMD	1	0,24%	0	0,00%	0	0,00%	0	0,00%
Collana	0	0,00%	0	0,00%	2	0,38%	0	0,00%
Collocazione	0	0,00%	1	0,11%	1	0,19%	0	0,00%
Data	16	3,87%	26	2,96%	30	5,67%	12	2,53%
DOI	0	0,00%	2	0,23%	5	0,95%	3	0,63%
Editore	4	0,97%	11	1,25%	7	1,32%	5	1,05%
Edizione	1	0,24%	1	0,11%	0	0,00%	1	0,21%

Formato	0	0,00%	0	0,00%	0	0,00%	1	0,21%
Identificativo_EDS	0	0,00%	5	0,57%	0	0,00%	0	0,00%
Identificativo_Catalogo Polo Bolognese	1	0,24%	0	0,00%	0	0,00%	0	0,00%
ISBN	1	0,24%	3	0,34%	2	0,38%	1	0,21%
ISSN	0	0,00%	0	0,00%	0	0,00%	2	0,42%
Lingua	0	0,00%	0	0,00%	0	0,00%	2	0,42%
Luogo	4	0,97%	8	0,91%	6	1,13%	1	0,21%
Natura_oggetto	0	0,00%	3	0,34%	1	0,19%	1	0,21%
Nome_monografia²⁰	3	0,73%	4	0,46%	0	0,00%	4	0,84%
Nome_periodico²¹	3	0,73%	4	0,46%	12	2,27%	9	1,89%
Pagine	2	0,48%	3	0,34%	9	1,70%	2	0,42%
Titolo	401	97,09%	848	96,58%	510	96,41%	464	97,68%
URL	0	0,00%	5	0,57%	0	0,00%	1	0,21%
Volume_fascicolo	6	1,45%	5	0,57%	14	2,65%	3	0,63%

Tabella 18 - *Elementi con cui è ricercato un documento conosciuto nelle quattro giornate analizzate*

²⁰ La voce fa riferimento alla monografia in cui è contenuto il capitolo cercato.

²¹ La voce fa riferimento al periodico in cui è contenuto l'articolo cercato.

Le ricerche mediante citazioni bibliografiche sono state rilevate perché spesso il discovery non è in grado di interpretarle correttamente²² e sono comprese tra il 5% e il 10% (tabella 19). Il dato più basso riguarda l'ultima giornata analizzata in cui sono state identificate 25 citazioni, numero che deve far riflettere sull'importanza di permettere agli utenti di effettuare ricerche mediante citazioni bibliografiche.

	Gior- no 1: 3 feb 20	% ri- spetto al numero ricerche docu- menti co- nosciti	Gior- no 2: 30 mar 20	% ri- spetto al numero ricerche docu- menti co- nosciti	Gior- no 3: 1 feb 21	% ri- spetto al numero ricerche docu- menti co- nosciti	Gior- no 4: 29 mar 21	% ri- spetto al numero ricerche docu- menti co- nosciti
Numero ricerche mediante citazioni	37	8,96%	47	5,35%	52	9,83%	25	5,26%

Tabella 19 - *Ricerche mediante citazioni bibliografiche e percentuale rispetto alle ricerche di documenti conosciuti*

L'analisi delle ricerche per argomento si è focalizzata sulla lingua. Capire quante ricerche siano fatte in una lingua diversa dall'inglese risulta interessante perché la maggior parte del materiale indicizzato nell'indice preconstituito di EDS è in lingua inglese.

Le ricerche per argomento in inglese sono sempre lievemente superiori a quelle in italiano, tranne che nel secondo bimestre 2020 (tabella 20). Probabilmente la scelta della lingua è in relazione al fatto che la letteratura scientifica sull'argomento cercato sia prevalentemente in inglese, ma il dato ottenuto potrebbe anche dipendere dal fatto che una parte degli utenti è consapevole che la maggior parte delle risorse

²² Va precisato che il discovery EDS offre la possibilità di attivare una funzione specifica per sciogliere le citazioni bibliografiche e che l'Università di Bologna ha deciso di non attivarla perché per utilizzarla l'utente dovrebbe selezionarla appositamente.

indicizzate nel discovery sono in lingua inglese e che quindi una ricerca in quella lingua può risultare più efficace.

Infatti nel maggio 2021 sulle 87 risorse bibliografiche italiane presenti nell'elenco delle collezioni elettroniche dell'Università di Bologna solo una risulta indicizzata all'interno di EDS. Si tratta della piattaforma editoriale Rivisteweb, contenente periodici di scienze umane e sociali pubblicati dagli editori Il Mulino e Carocci. Ovviamente alcuni contributi bibliografici italiani sono presenti anche in banche dati straniere, come ad esempio nella banca dati Web of Science indicizzata in EDS. Vale la pena riflettere però sul fatto che, in linea di massima, facendo una ricerca in italiano per argomento nel discovery si ottengono risorse principalmente dai cataloghi e repository locali, il Catalogo del Polo Bolognese, ACNP e AlmaDL, e quindi l'efficacia dello strumento risulta abbastanza circoscritta.

	Giorno 1: 3 feb 20	% rispetto a ricerche per argomento	Giorno 2: 30 mar 20	% rispetto a ricerche per argomento	Giorno 3: 1 feb 21	% rispetto a ricerche per argomento	Giorno 4: 29 mar 21	% rispetto a ricerche per argomento
Cinese²³	2	0,61%	0	0,00%	0	0,00%	0	0,00%
Francese	1	0,30%	5	1,21%	5	1,20%	1	0,21%
Greco	4	1,23%	0	0,00%	0	0,00%	0	0,00%
Inglese	150	46,29%	176	42,71%	200	48,19%	223	48,79%
Inglese_Italiano	0	0,00%	1	0,24%	1	0,24%	3	0,65%

²³ Le ricerche in cinese, greco e russo sono state effettuate nei caratteri propri della lingua.

Italiano	146	45,06%	208	50,48%	172	41,44%	182	39,82%
Latino	1	0,30%	0	0,00%	0	0,00%	2	0,43%
Portoghe- se	0	0,00%	0	0,00%	6	1,44%	0	0,00%
Russo	0	0,00%	0	0,00%	1	0,24%	0	0,00%
Spagnolo	0	0,00%	0	0,00%	0	0,00%	5	1,09%
Tedesco	0	0,00%	1	0,24%	0	0,00%	2	0,43%
Molteplici possibilità	16	4,93%	1	0,24%	28	6,74%	13	2,84%
Non deter- minabile	4	1,23%	20	4,85%	2	0,48%	26	5,68%
TOTALE	324		412		415		457	

Tabella 20 - *Lingua con cui sono state effettuate le ricerche per argomento nelle quattro giornate analizzate*

Durante l'analisi puntuale delle stringhe sono state identificate alcune ricerche che contengono elementi che difficilmente trovano una risposta adeguata mediante un discovery tool (tabella 21).

Queste ricerche, così come quelle che non hanno portato a nessun risultato nel discovery, danno informazioni estremamente rilevanti perché da un lato mostrano la scarsa consapevolezza di quali risorse metta a disposizione il discovery da parte di chi lo utilizza, dall'altro indicano possibili margini di miglioramento sulle funzionalità del discovery.

Ad esempio le ricerche per editore suggeriscono che sarebbe utile indicizzare tutti i database e le risorse bibliografiche presenti nel repertorio delle collezioni elettroniche dell'Università di Bologna, con tutti i dati relativi a esse compreso l'editore. Si potrebbero inoltre con-

siderare alcune parole di ricerca come filtri nella ricerca stessa e non come vere e proprie parole da trovare nel record: parole che indicano la natura dell'oggetto come "journal", "rivista" o "quotidiano" potrebbero essere considerate come un filtro per la natura periodico, mentre il formato "pdf" potrebbe essere considerato un filtro su risorse disponibili a testo pieno online.

Dal momento che sono presenti tre ricerche riguardanti AlmaStart si può supporre inoltre che il link alle informazioni sul discovery non sia così evidente.

Tipologia ricerca	Descrizione	Numero ricerche
Editore	Le ricerche per editore sono state messe in evidenza perché ci si è chiesti se si cercassero documenti di quell'editore oppure le piattaforme editoriali; nel secondo caso il discovery non è in grado di fornire una risposta, perché non vengono indicizzate	7
URL	Ricerche tramite URL di cataloghi o di risorse. Le URL dei DOI e i permalink delle risorse vengono indicizzate, ma le altre URL no	5
Natura oggetto	La specificazione della natura dell'oggetto cercato (journal, rivista, quotidiano) fa sì che le ricerca non porti all'oggetto desiderato. Ad esempio la stringa "space policy journal" non ha trovato il periodico "Space policy" nel momento in cui si è provato a replicare la ricerca	4
AlmaStart	Ricerca di informazioni su AlmaStart, disponibili sul sito del Sistema Bibliotecario e non all'interno del discovery stesso	3
Collocazione	La ricerca per collocazione funziona nel Catalogo del Polo Bolognese, mentre il dato non viene indicizzato nel discovery tool	2

Corso accademico	Due ricerche corrispondono esattamente a corsi universitari offerti dall'Università di Bologna, pertanto ci si è chiesti se si cercassero informazioni bibliografiche o informazioni sul corso accademico	2
Articolo quotidiano	È stato cercato un articolo comparso sul quotidiano La Stampa il 30 marzo 2020, stesso giorno in cui è stata fatta la ricerca. La ricerca non può trovare il risultato cercato non solo perché i quotidiani italiani non sono indicizzati dal discovery, ma anche perché l'indicizzazione dei documenti bibliografici non è contestuale alla loro pubblicazione, avviene dopo un certo lasso di tempo	1
Copia da Amazon	Ricerca che contiene gli elementi forniti generalmente dal sito di e-commerce Amazon, ovvero la data di pubblicazione con il giorno specifico e il formato "(copertina flessibile)". Sia il formato che la data esatta non sono presenti fra i dati bibliografici dei libri e pertanto non sono ricercabili	1
Formato	Alla ricerca è stato aggiunto il formato per il materiale, pdf. Il dato nel discovery non viene indicizzato, pertanto la ricerca "pascal pensieri pdf" probabilmente non ha portato al risultato desiderato	1
Materiale didattico	Il materiale didattico dell'Università di Bologna è su una piattaforma separata di e-learning e non viene indicizzato nel discovery	1
Servizio biblioteche	È stato cercato il servizio di prenotazione delle biblioteche. Le informazioni sono disponibili sul sito del Sistema Bibliotecario e non all'interno del discovery stesso	1
Traduzione di termine	È stata chiesta la traduzione di un termine dall'inglese all'italiano mediante la ricerca "voice over traduzione". Se questa viene effettuata sul motore di ricerca Google viene proposta la traduzione del termine, probabilmente l'utente si aspettava una funzionalità simile nel discovery	1

Tabella 21 - *Elementi non presenti nel discovery cercati nelle quattro giornate analizzate*

Altre informazioni interessanti che in parte confermano quanto già osservato emergono dall'analisi delle ricerche che non hanno prodotto risultati. È importante premettere che non è possibile determinare con certezza perché una ricerca non abbia prodotto risultati per due motivi: da un lato Google Analytics registra solo i filtri relativi ai cataloghi e non quelli di altro tipo (es: data, tipologia risorsa...), pertanto non è possibile sapere se la ricerca non ha portato risultati perché erano stati applicati dei filtri specifici. Inoltre le ricerche sono state replicate nel discovery nel mese di maggio 2021, ma sia il materiale indicizzato nel discovery che quello posseduto dall'Università di Bologna variano nel tempo.

Le ricerche per argomento senza risultati non sono state raggruppate né ai casi di mancata indicizzazione né a quelli dei documenti non disponibili perché non è possibile sapere se non sia posseduta alcuna pubblicazione su quell'argomento oppure non sia mai stata pubblicata (tabella 22).

Fra le ricerche che non hanno portato a risultati a causa della mancata indicizzazione è utile osservare che ci sono casi in cui lo specifico capitolo cercato non è indicizzato, ma è presente invece la monografia in cui è contenuto (etichetta Non indicizzato capitolo, c'è libro della tabella 22). Infatti questi casi mettono in risalto il problema della granularità dei record presenti nell'indice centrale dei discovery tool. Le collezioni di e-book possedute dall'Università di Bologna sono catalogate nel Catalogo del Polo Bolognese e quindi già rese disponibili anche tramite il discovery. Il valore aggiunto di AlmaStart consiste nel mettere a disposizione record che descrivono i singoli capitoli delle monografie e di record con abstract che riportano l'indice completo rendendolo ricercabile.

È utile specificare che nel Catalogo del Polo Bolognese vengono indicizzati gli indici e sommari di alcuni volumi posseduti dalle biblioteche che aderiscono a uno specifico progetto per metterli a disposizione, ma questi attualmente non vengono a loro volta indicizzati nel discovery perché i dati non sono esposti tramite protocollo

OAI-PMH e quindi il discovery non può importarli. Superare questo mancato colloquio tra i due strumenti arricchirebbe ulteriormente l'indice del discovery e permetterebbe agli utenti di reperire in Alma-Start esattamente tutto il materiale che trovano nel Catalogo del Polo Bolognese facendo la stessa ricerca.

In aggiunta ai documenti non posseduti e non indicizzati, si sono identificate un insieme di cause che hanno a che vedere con la ricerca effettuata dall'utente, con il funzionamento del discovery e con i campi presenti nei record. Fra queste le più numerose sono quelle riguardanti problemi di battitura. Normalmente i motori di ricerca mettono a disposizione la funzione "Forse cercavi", che permette all'utente di correggere immediatamente la ricerca nel caso abbia fatto errori di battitura. Anche EDS mette a disposizione questa funzionalità che è stata testata durante la fase di implementazione del discovery ma in seguito è stata disattivata perché, pur funzionando correttamente per la lingua inglese, non supportava invece la lingua italiana, proponendo termini in spagnolo per quasi tutte le ricerche in italiano. Indubbiamente mettere a disposizione questa funzione aiuterebbe gli utenti nella ricerca.

Anche la ricerca di citazioni è un problema abbastanza frequente, perché il discovery non è in grado di interpretarle e proporre il risultato corrispondente. Sicuramente i bibliotecari dovranno rivalutare la possibilità di introdurre una funzione separata per interpretarle correttamente, che viene messa a disposizione da EDS, in un primo momento scartata perché dalle prove effettuate non era comunque in grado di interpretare tutte le citazioni inserite.

Nel contempo si nota però che anche citazioni senza particolari abbreviazioni o segni di punteggiatura falliscono nella ricerca, per il fatto che in molti record le date non sono ricercabili ma possono solo essere usate come filtro. Ad esempio con la stringa "J. Weiler (ed), *The EU, the WTO and the NAFTA. Towards a Common Law of International Trade, 2001*" non si trovano risultati, mentre con la stessa stringa senza la data "J. Weiler (ed), *The EU, the WTO and the NAFTA.*

Towards a Common Law of International Trade,” il risultato desiderato appare al terzo posto.

Si nota infine una ricerca per formato (pdf), tipologia di ricerca già osservata fra le ricerche di elementi non presenti nel discovery. La ripetuta presenza di questo tipo di ricerca porta a chiedersi se fosse possibile tecnicamente fare in modo che l’inserimento della parola “pdf” si trasformi automaticamente in una ricerca che ometta quel termine e filtri per documenti disponibili online. Oltre all’aspetto tecnico, andrebbe preventivamente valutato se questa modifica potrebbe avere ripercussioni negative, aumentando il rumore dei risultati trovati.

	gen- feb 2020	mar-apr 2020	gen- feb 2021	mar-apr 2021	TO- TALE	%
Nessun risultato a causa della modalità di ricerca dell’utente, del funzionamento del discovery, dei record catalografici	45	40	34	76	195	42,95%
Battitura: mancanza di spazi fra le parole	5	6	3	13	27	
Battitura: errori di spelling	22	17	10	34	83	
Ricerca per citazione	12	13	14	18	57	
Ricerca di una collocazione	0	2	1	0	3	
Dati documento cercato non corretti	4	1	3	7	15	
Formato	0	0	0	1	1	
Inserito “EISBN:” prima dell’ISBN cercato	0	0	1	0	1	
Non indicizzato con identificativo usato per la ricerca	0	1	1	2	4	
Numeri romani e non arabi	0	0	1	0	1	
Ricerca tramite URL	2	0	0	1	3	

Mancata indicizzazione	22	14	14	24	74	16,30%
Ricerca documento conosciuto: documento non indicizzato	18	10	14	16	58	
Niente di indicizzato in ACNP o in AlmaDL (ricerca con filtri ACNP/AlmaDL)	0	2	0	3	5	
Non indicizzato capitolo, c'è libro	4	2	0	5	11	
Nessun documento disponibile (ricerca documento conosciuto)	30	22	20	26	98	21,59%
Mancata indicizzazione oppure nessun documento disponibile (ricerca per argomento)	3	6	2	17	28	6,17%
Problema non identificato	15	15	7	22	59	13,00%
TOTALE	115	97	77	165	454	

Tabella 22 - Possibili cause risultato nullo nei quattro bimestri analizzati

5. Conclusioni

Dai dati analizzati emerge che nel secondo bimestre del 2020 il lockdown ha avuto un impatto evidente sui dati di utilizzo del discovery, notevolmente incrementati, ma che le modalità di ricerca degli utenti sono rimaste abbastanza simili, rimanendo tali anche nell'anno successivo; ciò lascia supporre che i nuovi utenti che hanno iniziato a utilizzare il discovery durante il lockdown abbiano lo stesso livello di competenze informative di chi l'ha utilizzato nei bimestri precedenti e successivi.

Più del 92% delle ricerche partono dalla ricerca base e meno

del 11% utilizzano operatori booleani, caratteri speciali o la ricerca per campi specifici del record, come titolo o soggetto. Questo significa che la modalità utilizzata per 9 ricerche su 10 sia la stessa adottata sui motori di ricerca, un insieme di parole inserite in un unico campo.

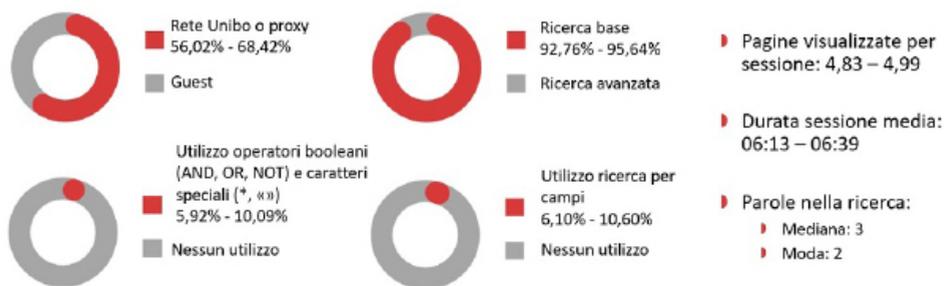


Figura 7 - Dati riassuntivi sulla modalità di ricerca nel discovery relativi ai quattro bimestri analizzati

Nel momento in cui appaiono i risultati, fra il 32% e il 34% delle ricerche porta alla visualizzazione del dettaglio di un record e circa il 35% delle ricerche porta all'utilizzo di un link, a eccezione del secondo bimestre del 2020 in cui vengono visualizzati meno del 29% dei record e utilizzati più del 46% dei link rispetto alle ricerche effettuate.

Questi dati portano a pensare che l'interesse principale degli utenti sia reperire il full text delle risorse, rispondendo quindi alla funzione "ottenere" identificata dal modello di rappresentazione dei documenti dell'IFLA Library Reference Model. Vengono utilizzate molto poco, invece, le funzioni presenti all'interno dei record che permettono di esportare o salvare i record bibliografici. Questo comportamento è ancor più accentuato nel secondo bimestre del 2020 ed è lo stesso che era emerso dall'analisi di utilizzo del discovery tool dell'Università di Siena.²⁴

La stessa ipotesi viene confermata dall'esame delle stringhe di ri-

²⁴ L'analisi effettuata sul discovery tool dell'Università di Siena, presentata in Machetti 2016, ha previsto l'analisi qualitativa dell'uso mediante l'osservazione degli utenti durante l'utilizzo dello strumento.

cerca: il discovery è utilizzato in primo luogo per cercare documenti di cui si conosce già l'esistenza e solo in secondo luogo per ricerche su un determinato argomento. Fra i documenti conosciuti cercati 1 su 4 è un libro e 1 su 10 un articolo.

Il fatto che una ricerca su quattro sia di una monografia e che le monografie, sia cartacee che online, siano tutte presenti all'interno del Catalogo del Polo Bolognese porta a domandarsi perché non venga utilizzato il catalogo al posto del discovery, dato che su questo previa autenticazione è possibile avere in aggiunta i servizi di richiesta o prenotazione delle copie cartacee. Una ipotesi plausibile è che il discovery venga utilizzato per cercare gli e-book e venga preferito rispetto al catalogo perché è più immediato loggarsi attraverso il proxy.²⁵ Si è visto infatti che fra il 56% e il 68% delle ricerche avviene da utenti che si sono autenticati.

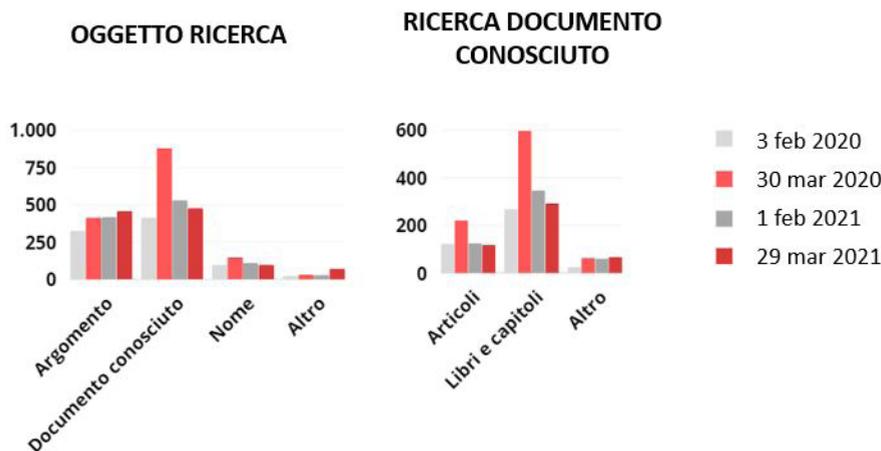


Figura 8 - Grafici sull'oggetto della ricerca effettuata e sulle tipologie di documento conosciuto cercato nelle quattro giornate analizzate

²⁵ Nel Catalogo del Polo Bolognese è prevista una funzione per l'autenticazione, ma è diversa da quella necessaria per il proxy. Per accedere a un e-book dal catalogo, se ci si trova connessi a una rete esterna a quella dell'Università di Bologna, è necessario prima autenticarsi con il proxy, poi arrivare al catalogo ed effettuare la ricerca.

Dai dati riguardanti la ricerca di informazioni non presenti nel discovery e dalle ricerche che non hanno portato ad alcun risultato si ottengono informazioni utili sulle aspettative degli utenti riguardo al funzionamento e al contenuto dello strumento che suggeriscono una serie di elementi da tenere in considerazione per migliorarne le funzionalità, sia per i bibliotecari che si occupano delle configurazioni di AlmaStart che per il produttore del discovery.

È emerso che molte ricerche non producono risultati a causa di problemi di battitura, che al momento producono un risultato nullo, senza offrire la possibilità di una ricerca alternativa come fanno usualmente i motori di ricerca sul web, mentre sarebbe opportuno che la funzionalità “Forse cercavi” venisse resa disponibile anche per la lingua italiana oltre che per la lingua spagnola e inglese.

Vengono usate citazioni bibliografiche per effettuare le ricerche ma queste di solito non producono risultati a causa delle abbreviazioni e dei molteplici elementi che le compongono; altro elemento che crea difficoltà nel reperimento dei risultati è la data, che nella maggior parte dei record non è un elemento ricercabile, mentre sarebbe auspicabile che lo fosse.

Si rileva l'abitudine degli utenti di inserire nel campo di ricerca elementi che generalmente nei cataloghi bibliografici sono considerati filtri, come la natura dell'oggetto o il formato di questo. Si potrebbe valutare la possibilità di convertire alcuni elementi della ricerca in filtri e di non considerare i termini relativi alla natura e al formato del materiale come parole da cercare.

Sono state individuate infine un insieme di ricerche che contengono esclusivamente il nome dell'editore, da cui si suppone che l'utente volesse identificare una banca dati o l'insieme delle risorse messe a disposizione dall'editore stesso; sarebbe quindi importante indicizzare fra le risorse locali anche il repertorio delle collezioni sottoscritte dall'Università di Bologna.

Un ultimo elemento chiave, che spesso viene dato per scontato ma su cui vale la pena di soffermarsi a riflettere, è il grado di copertura

dell'indice centrale del discovery rispetto alle collezioni sottoscritte dall'istituzione. È evidente che se si mette a disposizione uno strumento con lo scopo di permettere una ricerca unificata ma poi questo cerca in un sottoinsieme molto limitato di risorse, fallisce il suo scopo principale. Dalla ricerca è emerso che l'italiano è la lingua utilizzata per circa il 40% delle ricerche per argomento. Ritengo quindi importante che i bibliotecari insistano con gli editori italiani con cui stipulano contratti affinché forniscano i propri metadati alle aziende che producono i discovery tools; questo va a vantaggio della propria utenza finale ma anche degli editori stessi, dato che la conseguenza è un maggior utilizzo delle risorse che diventano così più essenziali per le università che le sottoscrivono. Allo stesso tempo i bibliotecari possono sollecitare i produttori di discovery a indicizzare prodotti del mercato editoriale italiano, suggerendo quelli delle aree disciplinari per cui lo strumento viene più utilizzato, come le scienze umane e sociali, e più in generale quelle che si ritiene essere più rilevanti per la propria utenza.

Bibliografia

- Bardelli 2019 = Annalisa Bardelli, Francesca Verga, *Il Discovery tool è un organismo in crescita*, «JLIS.it», 10 (2019), 3, p. 97-124.
- Bianchini 2017 = Carlo Bianchini, «*Funziona come Google, vero?*». *Prima indagine sull'interazione utente-catalogo nella biblioteca del Dipartimento di musicologia e beni culturali (Cremona) dell'Università di Pavia*, «AIB studi», 57 (2017), 1.
- Bianchini 2020 = Carlo Bianchini, «*Andrò a cercare tra gli scaffali*»: *indagine conoscitiva sull'interazione lettore-catalogo nella Biblioteca Civica "Vincenzo Joppi" di Udine*, «Bibliothecae.it», (2020), p. 420-457.
- Cohen 2015 = Rachael A. Cohen, Angie Thorpe Pusnik, *Discovering User Behavior: Applying Usage Statistics to Shape Frontline Services*, «The Serials Librarian», 69 (2015), 1, p. 29-46.
- Fischer 2021 = Rachel K. Fischer, Aubrey Iglesias, Alice L. Daugherty, [et al.], *A transaction log analysis of EBSCO Discovery Service using Google Analytics: the methodology*, «Library Hi Tech», 39 (2021), 1, p. 249-262.
- Francesco 2013 = Enrico Francesco, *Test di usabilità sul discovery tool «Primo» all'Università di Torino*, «Biblioteche Oggi», 31 (2013), 10, p. 10-17.
- Friese 2020 = Susanne Friese, *Creating a Coding Scheme with ATLAS.ti*, ATLAS.ti, <https://atlasti.com/2020/01/31/creating-a-coding-scheme-with-atlas-ti-by-susanne-friese/> (18.05.2021).
- Fusetti 2019 = Camilla Fusetti, Enrico Tagliani, *Analizzare per innovare: uno studio sull'usabilità dei cataloghi Discovery NG con strumenti di web analytics*, «Biblioteche oggi Trends», 5 (2019), 1, p. 26-36.
- Galeffi 2017 = Agnese Galeffi, *Se il catalogo parlasse, lo capiremmo? Cinque assiomi della comunicazione catalografica*, «AIB studi», 57 (2017), 2.
- Machetti 2016 = Clarissa Machetti, *Biblioteche e discovery tool: il caso One-Search e l'ateneo di Siena*, «AIB studi», 56 (2016), 3, p. 391-408.
- Marek 2011 = Kate Marek, *Using web analytics in the library*, Chicago, Ill., ALA TechSource, 2011.

Bibliothecae.it
11 (2022), 1, 358-360
Saggi

Maura Quaquarelli
*Utilizzo del discovery tool dell'Università di Bologna:
un'analisi quantitativa e qualitativa*

Milesi 2011 = Patrizia Milesi, Patrizia Cattelani, *L'analisi qualitativa dei testi con il programma Atlas.ti*, in *Metodi qualitativi in psicologia sociale: prospettive teoriche e strumenti operativi*, Roma, Carocci, 2011.

Abstract

La ricerca ha lo scopo di identificare le modalità di ricerca più comuni adottate dagli utenti in un discovery tool e se questo sia in grado di dare risposte adeguate a chi lo utilizza. Lo studio viene condotto sull'installazione di EBSCO Discovery Service dell'Università di Bologna, AlmaStart. Vengono analizzati i dati relativi a quattro bimestri, i primi due bimestri del 2020 e del 2021, che permettono di verificare se le modalità di ricerca cambino nel tempo e se il lockdown del 2020, dovuto alla pandemia da Coronavirus, abbia modificato i comportamenti di ricerca. L'uso del discovery tool dell'Università di Bologna viene analizzato partendo dai dati sulle ricerche raccolti da Google Analytics, mediante l'utilizzo dei software Microsoft Excel e Atlas.ti, un software che consente l'analisi e la categorizzazione di dati testuali. Viene fatta un'analisi quantitativa sui dati relativi a quattro bimestri e un'analisi qualitativa sulle stringhe di ricerca relativa a quattro specifiche giornate e a tutte le ricerche che non hanno prodotto alcun risultato. L'analisi qualitativa ha reso possibile capire se gli utenti cerchino prevalentemente documenti di cui conoscono già l'esistenza oppure usino il discovery per ricerche su un particolare argomento; si sono inoltre intercettate alcune aspettative dell'utenza sul funzionamento della ricerca che il discovery non riesce a soddisfare e quali siano i punti di attenzione per il miglioramento dello strumento. Si è scoperto che l'utilizzo dello strumento durante il lockdown ha subito un incremento consistente, ma che le modalità di ricerca adottate non hanno subito particolari cambiamenti. Il discovery viene utilizzato principalmente per trovare documenti di cui si conosce già l'esistenza con lo scopo di raggiungere il full text della risorsa. Si sono identificati quindi una serie di elementi utili a migliorarne la funzionalità.

Discovery tool; Università di Bologna; Comportamento di ricerca; Atlas.ti

This article has the purpose to identify the most common search techniques of patrons in a discovery tool and if the discovery is able to provide relevant answers to the queries. This study reports on the use of EBSCO Discovery Service of Bologna University, called AlmaStart. The analyzed data refers to 8 months, from January to April 2020 and from January to April 2021. This period allowed to verify if search behavior changed over time and if the closure of the libraries due to coronavirus lockdown affected it. The data, collected by Google Analytics, are analyzed through the use of the software Microsoft Excel and Atlas.ti, a tool for the qualitative analysis and categorization of textual data. The quantitative analysis refers to all collected data, the qualitative analysis refers to 4 specific days and to all search queries that did not conduct to any results in the discovery. The qualitative analysis allowed to verify if patrons mostly conduct searches about a theme, or if they use the discovery to search known-items in order to access or to localize them. Moreover, it was possible to identify some patron's expectations about search queries that discovery is not able to satisfy and some ideas to improve it. The result reveals that during the library closure the use of the discovery tool increased a lot, but the search behaviour remained basically the same as in the other months. Patrons use the discovery mostly to search known-item in order to reach the full text of the documents. Finally, some elements to improve the discovery tool functionality were identified.

Discovery tools; University of Bologna; Searching behavior; Atlas.ti